Joint Posterior Revision of NLP Annotations via Ontological Knowledge

Marco Rospocher
Francesco Corcoglioniti
Kia has hired Peter Schreyer as chief design officer.
Kia has hired Peter Schreyer as chief design officer.
Context: Knowledge Extraction

Kia has hired Peter Schreyer as chief design officer.

NLP Tasks:
- Named Entity Recognition and Classification (NERC)
Context: Knowledge Extraction

Kia has hired Peter Schreyer as chief design officer.

NLP Tasks:
- Named Entity Recognition and Classification (NERC)
- Entity Linking (EL)
Kia has hired Peter Schreyer as chief design officer.

NLP Tasks:
- Named Entity Recognition and Classification (NERC)
- Entity Linking (EL)
- Semantic Role Labeling (SRL)
...
Motivating Examples

Mr. Washington was runner-up at Wimbledon in 1996.
Motivating Examples

Mr. Washington was runner-up at Wimbledon in 1996.

Stanford CoreNLP
http://nlp.stanford.edu:8080/corenlp
Motivating Examples

Mr. Washington was runner-up at Wimbledon in 1996.
**Motivating Examples**

Mr. **Washington** was runner-up at Wimbledon in 1996.

<table>
<thead>
<tr>
<th>Stanford CoreNLP</th>
<th>DBpedia Spotlight</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="http://nlp.stanford.edu:8080/corenlp" alt="Person" /></td>
<td><img src="http://demo.dbpedia-spotlight.org" alt="DBpedia Spotlight" /></td>
</tr>
</tbody>
</table>

The **GW Bridge** is a double-decked suspension bridge over the Hudson.
Motivating Examples

Mr. Washington was runner-up at Wimbledon in 1996.

The GW Bridge is a double-decked suspension bridge over the Hudson.
Motivating Examples

Mr. Washington was runner-up at Wimbledon in 1996.

The GW Bridge is a double-decked suspension bridge over the Hudson.
Abstracting

... token₁ token₂ token₃ token₄ token₅ token₆ ....
Abstracting

… token\textsubscript{1} token\textsubscript{2} \textbf{token\textsubscript{3} token\textsubscript{4}} token\textsubscript{5} token\textsubscript{6} ….
Abstracting

\[ \text{Task}_1 \quad \text{Task}_2 \quad \text{Task}_n \]

\[ \ldots \text{token}_1 \quad \text{token}_2 \quad \text{token}_3 \quad \text{token}_4 \quad \text{token}_5 \quad \text{token}_6 \quad \ldots \]
Abstracting

\[
\begin{align*}
\begin{array}{c}
\text{Task}_1 \\
\vdots \\
\text{Task}_k \\
\end{array}
\end{align*}
\]

\[
\begin{align*}
\begin{array}{c}
\text{Task}_1 \\
\vdots \\
\text{Task}_k \\
\end{array}
\end{align*}
\]

\[
\begin{align*}
\begin{array}{c}
\text{Task}_1 \\
\vdots \\
\text{Task}_k \\
\end{array}
\end{align*}
\]

... token_1 \ token_2 \ token_3 \ token_4 \ token_5 \ token_6 \ ...
Abstracting

... token_1 token_2 token_3 token_4 token_5 token_6 ....
Abstracting

... token_1 token_2 token_3 token_4 token_5 token_6 ....
RESEARCH PROBLEM

How can we assess and improve the coherence of the various NLP annotations on an entity mention?
In a nutshell

ontological background knowledge

\[ a_{1,1}, a_{1,2}, \ldots, a_{1,k}, a_{2,1}, a_{2,2}, \ldots, a_{2,i}, \ldots, a_{n,1}, a_{n,2}, \ldots, a_{n,j} \]

\[ \ldots \text{token}_1 \text{token}_2 \text{token}_3 \text{token}_4 \text{token}_5 \text{token}_6 \ldots \]
In a nutshell

ontological background knowledge

Task_1

\[
\begin{align*}
& a_{1,1} \quad a_{1,2} \\
& \vdots \\
& a_{1,k}
\end{align*}
\]

Task_2

\[
\begin{align*}
& a_{2,1} \quad a_{2,2} \\
& \vdots \\
& a_{2,i}
\end{align*}
\]

Task_n

\[
\begin{align*}
& a_{n,1} \quad a_{n,2} \\
& \vdots \\
& a_{n,j}
\end{align*}
\]

... token_1 token_2 token_3 token_4 token_5 token_6 ...
In a nutshell
ontological background knowledge

\[ \begin{align*}
\text{Task}_1: a_{1,1}, a_{1,2}, \ldots, a_{1,k} \\
\text{Task}_2: a_{2,1}, a_{2,2}, \ldots, a_{2,i} \\
\text{Task}_n: a_{n,1}, a_{n,2}, \ldots, a_{n,j} \\
\end{align*} \]

\[ \begin{align*}
\text{Token}_1, \text{Token}_2, \text{Token}_3, \text{Token}_4, \text{Token}_5, \text{Token}_6, \ldots
\end{align*} \]
In a nutshell

ontological background knowledge

Task\textsubscript{1}

\[ a_{1,1}, a_{1,2}, a_{1,k} \]

Task\textsubscript{2}

\[ a_{2,1}, a_{2,2}, a_{2,i} \]

Task\textsubscript{n}

\[ a_{n,1}, a_{n,2}, a_{n,j} \]

\[ \cdots \]

\[ \text{... token}_1, \text{token}_2, \text{token}_3, \text{token}_4, \text{token}_5, \text{token}_6, \cdots \]
Contributions

1. **JPARK**: a probabilistic model capable to estimate a posteriori the overall confidence of NLP annotations

2. A concrete instantiation of the model for NERC and EL (using YAGO as ontological knowledge)

3. Application of the NERC and EL model to revise the annotations of Stanford NER and DBpedia Spotlight
The JPARK Model

\[ P(\alpha | m, B, K) \]
The JPARK Model

entity mention

\((a_i, \ldots, a_n)\) NLP Annotations

NLP Background Knowledge

“The” Ontological Knowledge

\(P(\mathbf{a} | m, B, K)\)
The JPARK Model

entity mention \( (a_i, \ldots, a_n) \) NLP Annotations

NLP Background Knowledge

"The" Ontological Knowledge

\[
P(a \mid m, B, K)
\]

set of classes from \( K \)

\[
P(a, C \mid m, B, K)
\]
The JPARK Model

entity mention \( (a_i, \ldots, a_n) \) NLP Annotations

NLP Background Knowledge

“The” Ontological Knowledge

\[
P(\mathbf{a} \mid m, B, K)
\]

set of classes from \( K \)

\[
P(\mathbf{a}, \mathbf{C} \mid m, B, K)
\]

\[
P(a_i \mid m, B)
\]

\[
P(C \mid a_i, K)
\]
The **JPARK** Model

entity mention \( (a_i, \ldots, a_n) \) NLP Annotations

NLP Background Knowledge

P \( (a | m, B, K) \)

“The” Ontological Knowledge

set of classes from \( K \)

P \( (a, C | m, B, K) \)

confidence score

P \( (a_i | m, B) \) \[ \text{boxed} \]

P \( (C | a_i, K) \)
The Jpark Model

entity mention

NLP Annotations

(a₁, …, aₙ)

NLP Background Knowledge

“The” Ontological Knowledge

\[ P(\alpha|m, B, K) \]

set of classes from \( K \)

\[ P(\alpha, C|m, B, K) \]

confidence score

\[ P(a_i|m, B) \]

learned from data

\[ P(C|a_i, K) \]
The **JPARK** Model

$$P(a, C | m, B, K) = \arg \max_a P(a | m, B, K)$$

entity mention
NLP Background Knowledge
"The" Ontological Knowledge

set of classes from $K$

confidence score
learned from data

$$P(a_i | m, B)$$
$$P(C | a_i, K)$$
NERC and EL Model
Ingredients

• Ontological Knowledge

• Estimating $P (C | a_{NERC}, K)$

• Estimating $P (C | a_{EL}, K)$
Ingredients

• Ontological Knowledge

• Estimating $P (C | a_{\text{NERC}}, K)$

• Estimating $P (C | a_{\text{EL}}, K)$
Ingredients

• Ontological Knowledge

• Estimating \( P(C|a_{\text{NERC}}, K) \)

  Leverage a gold standard corpus \( G \) annotated with NERC types and ontological classes (or EL annotations)

• Estimating \( P(C|a_{\text{EL}}, K) \)
Ingredients

• Ontological Knowledge

• Estimating $P(C|a_{\text{NERC}}, K)$ ~ \( \frac{n_G(C, a_{\text{NERC}})}{\sum_C n_G(C, a_{\text{NERC}})} \)

Leverage a gold standard corpus $G$ annotated with NERC types and ontological classes (or EL annotations)

• Estimating $P(C|a_{\text{EL}}, K)$
Ingredients

- **Ontological Knowledge**

- **Estimating** \( P(C|a_{\text{NERC}}, K) \) \( \approx \frac{n_G(C, a_{\text{NERC}})}{\sum_C n_G(C, a_{\text{NERC}})} \)

  Leverage a **gold standard corpus** \( G \) annotated with NERC types and ontological classes (or EL annotations)

- **Estimating** \( P(C|a_{\text{EL}}, K) \)

  Leverage **alignments** between EL Knowledge Base and
Ingredients

- **Ontological Knowledge**

- **Estimating** \( P(C|a_{\text{NERC}}, K) \) \( \mathop{\sim} \frac{n_G(C, a_{\text{NERC}})}{\sum_C n_G(C, a_{\text{NERC}})} \)

Leverage a **gold standard corpus** \( G \) annotated with NERC types and ontological classes (or EL annotations)

- **Estimating** \( P(C|a_{\text{EL}}, K) \) \( \begin{cases} 1 & \text{entity } a_{\text{EL}} \text{ is instance of } C \\ 0 & \text{otherwise} \end{cases} \)

Leverage **alignments** between EL Knowledge Base and...
Application and Evaluation
Tools

• **NERC**: Stanford CoreNLP [Finkel et al., 2005]

• **EL**: DBpedia Spotlight [Daiber et al., 2013]
NERC+EL Datasets

- AIDA CoNLL-YAGO [Hoffart et al., 2011]
- MEANTIME [Minard et al., 2016]
- TAC-KBP [Ji et al., 2011]
Research Question

Does the JPARK posteriori joint revision of the annotations from Stanford NER and DBpedia Spotlight, via YAGO, improve their NERC and EL performances?
Research Question

Does the *JPARK* posteriori joint revision of the annotations from Stanford NER and DBpedia Spotlight, via YAGO, *improve* their NERC and EL performances?
## Results

<table>
<thead>
<tr>
<th></th>
<th>NERC</th>
<th>EL</th>
<th>NERC+EL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(P)</td>
<td>(R)</td>
<td>(F_1)</td>
</tr>
<tr>
<td><strong>AIDA</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>standard</em></td>
<td>94.30%</td>
<td>87.50%</td>
<td>90.80%</td>
</tr>
<tr>
<td><em>with JPARK</em></td>
<td>95.00%</td>
<td>88.10%</td>
<td>91.40%</td>
</tr>
<tr>
<td>(\Delta)</td>
<td>0.70%</td>
<td>0.60%</td>
<td>0.60%</td>
</tr>
<tr>
<td><strong>MEANITIME</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>standard</em></td>
<td>88.20%</td>
<td>69.50%</td>
<td>77.70%</td>
</tr>
<tr>
<td><em>with JPARK</em></td>
<td>91.40%</td>
<td>72.00%</td>
<td>80.50%</td>
</tr>
<tr>
<td>(\Delta)</td>
<td>3.20%</td>
<td>2.50%</td>
<td>2.80%</td>
</tr>
<tr>
<td><strong>TAC-KBP</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>standard</em></td>
<td>91.10%</td>
<td>65.20%</td>
<td>76.00%</td>
</tr>
<tr>
<td><em>with JPARK</em></td>
<td>92.60%</td>
<td>66.30%</td>
<td>77.20%</td>
</tr>
<tr>
<td>(\Delta)</td>
<td>1.50%</td>
<td>1.10%</td>
<td>1.20%</td>
</tr>
</tbody>
</table>

Bold = statistical significant (approx. rand. test)
## Results

<table>
<thead>
<tr>
<th>NER C</th>
<th>EL</th>
<th>NERC+EL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$P$</td>
<td>$R$</td>
</tr>
<tr>
<td><strong>AIDA</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>standard</td>
<td>94.30%</td>
<td>87.50%</td>
</tr>
<tr>
<td>with JPARK</td>
<td>95.00%</td>
<td>88.10%</td>
</tr>
<tr>
<td><strong>Δ</strong></td>
<td>0.70%</td>
<td>0.60%</td>
</tr>
<tr>
<td><strong>MEANTIME</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>standard</td>
<td>88.20%</td>
<td>69.50%</td>
</tr>
<tr>
<td>with JPARK</td>
<td>91.40%</td>
<td>72.00%</td>
</tr>
<tr>
<td><strong>Δ</strong></td>
<td>3.20%</td>
<td>2.50%</td>
</tr>
<tr>
<td><strong>TAC-KBP</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>standard</td>
<td>91.10%</td>
<td>65.20%</td>
</tr>
<tr>
<td>with JPARK</td>
<td>92.60%</td>
<td>66.30%</td>
</tr>
<tr>
<td><strong>Δ</strong></td>
<td>1.50%</td>
<td>1.10%</td>
</tr>
</tbody>
</table>

Bold = statistical significant (approx. rand. test)
## Results

<table>
<thead>
<tr>
<th></th>
<th>NERC</th>
<th>EL</th>
<th>NERC+EL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$P$</td>
<td>$R$</td>
<td>$F_1$</td>
</tr>
<tr>
<td><strong>AIDA</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>standard</strong></td>
<td>94.30%</td>
<td>87.50%</td>
<td>90.80%</td>
</tr>
<tr>
<td><strong>with JPARK</strong></td>
<td>95.00%</td>
<td>88.10%</td>
<td>91.40%</td>
</tr>
<tr>
<td><strong>$\Delta$</strong></td>
<td>0.70%</td>
<td>0.60%</td>
<td>0.60%</td>
</tr>
<tr>
<td><strong>MEANETIME</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>standard</strong></td>
<td>88.20%</td>
<td>69.50%</td>
<td>77.70%</td>
</tr>
<tr>
<td><strong>with JPARK</strong></td>
<td>91.40%</td>
<td>72.00%</td>
<td>80.50%</td>
</tr>
<tr>
<td><strong>$\Delta$</strong></td>
<td>3.20%</td>
<td>2.50%</td>
<td>2.80%</td>
</tr>
<tr>
<td><strong>TAC-KBP</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>standard</strong></td>
<td>91.10%</td>
<td>65.20%</td>
<td>76.00%</td>
</tr>
<tr>
<td><strong>with JPARK</strong></td>
<td>92.60%</td>
<td>66.30%</td>
<td>77.20%</td>
</tr>
<tr>
<td><strong>$\Delta$</strong></td>
<td>1.50%</td>
<td>1.10%</td>
<td>1.20%</td>
</tr>
</tbody>
</table>

Bold = statistical significant (approx. rand. test)
Research Question

Does the JPARK posteriori joint revision of the annotations from Stanford NER and DBpedia Spotlight, via YAGO, improve their NERC and EL performances?
Conclusions

• **Novel** probabilistic model, leveraging **ontological knowledge**, for improving NLP entity annotations

• Instantiation of the model for the **NERC** and **EL** tasks

• **Empirical confirmation** (3 datasets) of the capability of the model to improve the quality of the annotations

• **Future Work**: **extension** to other tasks (e.g., SRL)